

The Case for an Early Biological Origin of DNA

Anthony M. Poole · Nobuyuki Horinouchi ·
Ryan J. Catchpole · Dayong Si · Makoto Hibi ·
Koichi Tanaka · Jun Ogawa

Received: 20 October 2014 / Accepted: 18 November 2014 / Published online: 26 November 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract All life generates deoxyribonucleotides, the building blocks of DNA, via ribonucleotide reductases (RNRs). The complexity of this reaction suggests it did not evolve until well after the advent of templated protein synthesis, which in turn suggests DNA evolved later than both RNA and templated protein synthesis. However, deoxyribonucleotides may have first been synthesised via an alternative, chemically simpler route—the reversal of the deoxyriboaldolase (DERA) step in deoxyribonucleotide salvage. In light of recent work demonstrating that this reaction can drive synthesis of deoxyribonucleosides, we consider what pressures early adoption of this pathway would have placed on cell metabolism. This in turn provides a rationale for the replacement of DERA-dependent DNA production by RNR-dependent production.

Keywords Ribonucleotide reductase · Deoxyriboaldolase · Last Universal Common Ancestor · RNA world · DNA origins

Introduction

The RNA world hypothesis posits that templated protein synthesis and DNA-based storage of genetic information evolved well after RNA was already established as a major catalyst and as the genetic material (Gilbert 1986; Yarus 2011). Chemical (“bottom-up”) efforts to understand the RNA world have centred on the path from prebiotic chemistry to a simple RNA world (Anastasi et al. 2007; Benner et al. 1989; Dworkin et al. 2003; Hud et al. 2013) as well as ascertaining whether the catalytic capabilities of RNA could support the complex chemical metabolism presumed to be central to an RNA-based stage of life (Yarus 2002). In contrast, biological research has focused on whether the evolutionary history of RNA molecules is consistent with their presence early in the history of life (Hoeppner et al. 2012; Jeffares et al. 1998; White 1976). Likewise, a key part of this “top-down” or biological approach to testing the RNA world hypothesis has been to understand the nature of the evolutionary transitions from RNA to protein and from RNA to DNA. That the ribosome contains universally conserved RNAs, integrally involved in decoding and peptidyl transfer (Nissen et al. 2000; Noller 2010; Petrov et al. 2014), provides compelling evidence in favour of the RNA world, as does the demonstration that RNase P, a universal enzyme (Collins et al. 2000) required for tRNA maturation, is a catalytic RNA (Guerrier-Takada et al. 1983).

Equally, that all life synthesizes deoxyribonucleotides from ribonucleotide precursors via the ubiquitous (Lundin

Electronic supplementary material The online version of this article (doi:10.1007/s00239-014-9656-6) contains supplementary material, which is available to authorized users.

A. M. Poole · R. J. Catchpole
Biomolecular Interaction Centre, School of Biological Sciences,
University of Canterbury, Christchurch 8140, New Zealand

A. M. Poole (✉)
Allan Wilson Centre, University of Canterbury,
Christchurch 8140, New Zealand
e-mail: anthony.poole@canterbury.ac.nz

N. Horinouchi · D. Si · M. Hibi · K. Tanaka · J. Ogawa
Division of Applied Life Sciences, Graduate School of
Agriculture, Kyoto University, Kitashirakawa-Oiwakecho,
Sakyo-ku, Kyoto 606-8502, Japan

D. Si
Department of Molecular Bioscience and Bioengineering,
University of Hawaii, Honolulu, HI 96822, USA

et al. 2009), structurally homologous (Larsson et al. 2004; Logan et al. 1999; Sintchak et al. 2002; Tauer and Benner 1997; Uhlin and Eklund 1994) catalytic subunit of ribonucleotide reductases (RNRs), suggests that deoxyribonucleotide synthesis was added onto an already established ribonucleotide synthesis pathway (Lazcano et al. 1988; Poole et al. 2001, 2002; Reichard 1993; Torrents et al. 2002). However, ribonucleotide reduction is chemically complex, requiring the generation and containment of free radicals (Hofer et al. 2012). This requirement for free-radical chemistry suggests that the ribonucleotide reduction reaction, as carried out by modern RNRs, was beyond the capacity of RNA enzymes because RNA is indiscriminately cleaved by free radicals (Poole et al. 2000). This in turn suggests the evolutionary transition from RNA to DNA was preceded by the emergence of genetically encoded protein catalysts (Freeland et al. 1999; Poole et al. 2000) (Fig. 1).

Whereas biologists primarily see evolution of ribonucleotide reduction as the main roadblock to evolution of DNA (Freeland et al. 1999; Poole et al. 2000), deoxyribonucleotide synthesis is considered from a chemical perspective to be relatively straightforward compared with evolving a ribosome (Benner et al. 1989; Burton and Lehman 2009). In contrast to the possibility of an early chemical origin for DNA, comparative genomics analyses can be interpreted as compatible with a late origin for DNA (Forterre 1999; Forterre 2002; Leipe et al. 1999), perhaps post-dating the divergence of modern cells from a common ancestor (often dubbed LUCA, for Last Universal Common Ancestor). That said, comparative data are also compatible with an earlier, pre-LUCA origin for DNA (Poole and Logan 2005). Consequently, top-down and bottom-up

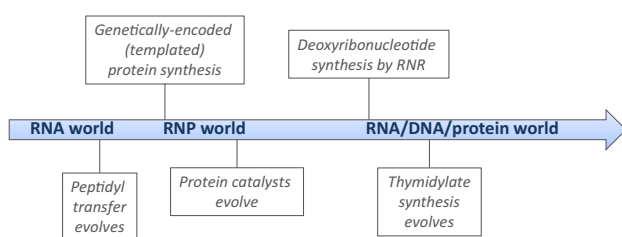


Fig. 1 Overview of the evolutionary transitions from RNA to protein and DNA. One view of the transitions from the RNA world to the modern DNA + protein world holds that the first transition coincided with the origin of templated protein synthesis (Noller 2010). The transition to DNA as genetic material is thought to have occurred later, requiring the evolution of genetically encoded protein enzymes. This view is based on the catalytic complexity of ribonucleotide reduction, which places the modern reaction beyond the capability of catalytic RNA (Poole et al. 2000). The transition from RNA to DNA is thought to be a two-stage transition, with replacement of the fourth base being a later step, possibly driven by futile cycling of deamination repair (Poole et al. 2001), or a virus-driven coevolutionary arms race (Forterre 2002)

views have diverged significantly on this point. The overarching goal of origins research is to build a coherent picture spanning prebiotic chemistry through to early evolution. To this end, it is worth considering whether an early case for DNA can be made from the biological ‘top-down’ perspective, and how such a model might be tested. We first review the apparently opposing views that have emerged for both the top-down and bottom-up models, then we turn to whether they might converge on a consensus.

A Top-Down Perspective on the Origin of DNA

While ribonucleotide reduction provides the sole mechanism for *de novo* synthesis of deoxyribonucleotides, the evolutionary history of RNRs, and even the DNA replication machinery, is complex. Consequently, it is difficult to place any complete DNA-associated processes in the LUCA (Forterre et al. 2004; Harris et al. 2003; Leipe et al. 1999). In stark contrast to ribosomal RNAs and many ribosomal proteins, which show a primarily vertical evolutionary history (Goldman et al. 2013; Harris et al. 2003; Woese and Fox 1977), the early evolutionary history of RNRs is peppered by interdomain horizontal transfer (Lundin et al. 2010), and there is too little sequence similarity to build sequence-based phylogenies spanning all three classes of RNRs (Lundin et al. 2010; Tauer and Benner 1997). The varying operational constraints of RNRs provide a clear explanation for this observation; RNRs have diversified into strictly anaerobic (class III), B12-dependent (class II, which operate irrespective of oxygen presence/absence) and two strictly aerobic forms (classes Ia & Ib) (Lundin et al. 2010). These classes are utilised in different environmental contexts (del Mar Cendra et al. 2012), and many microbes carry multiple classes of RNR (Lundin et al. 2009). Consequently, environment and horizontal transfer are key drivers of RNR inheritance patterns (Dwivedi et al. 2013; Lundin et al. 2010).

The other key enzymatic reaction central to the RNA to DNA transition is thymidylate synthesis (Fig. 1b). Thymidylate synthases, which convert dUMP to dTTP, were thought to have a single origin, but the discovery of a second class of thymidylate synthase (ThyX), unrelated to conventional thymidylate synthases (Myllykallio et al. 2002), shows that, while the reactions might be universal, the enzymes are not. The observation that the thymidylate synthase genes have distinct distributions, with the new class spread across a number of lineages, can best be understood in light of horizontal gene transfer (Myllykallio et al. 2002).

A number of key parts of the DNA replication machinery exhibit similarly complex histories. Most notably, replicative DNA polymerases and primases in bacteria are unrelated to those in archaea and eukaryotes (Augustin

et al. 2001; Bailey et al. 2006; Leipe et al. 1999), raising the possibility that modern processes for DNA synthesis and replication could have evolved after the primary bifurcation that gave rise to Archaea/Eukarya and Bacteria. In contrast, the conserved core of cellular RNA polymerases have a single evolutionary origin, and their evolutionary history appears broadly congruent with that of the ribosome (Cramer 2002; Tourasse and Gouy 1999; Werner and Grohmann 2011). A further series of observations that are consistent with a late origin for DNA are that modern RNA polymerases are capable of proofreading and repair and can replicate RNA templates (Lehmann et al. 2007; Poole and Logan 2005; Zenkin et al. 2006), notably in replication of plant viroids and Hepatitis Delta Virus (Fels et al. 2001; Lai 2005). This, coupled with an appreciation of the robustness of RNA to mutation (Kun et al. 2005), and the expected early origin of recombination (Lehman 2003; Reaney 1987), raises the possibility of large RNA genomes in early cellular lineages (Poole 2006; Poole and Logan 2005). In addition, accessory proteins which augment proofreading by RNA polymerase also evolved independently in Archaea/Eukarya and Bacteria after this primary diversification, suggesting that improvements to RNA fidelity were ongoing prior to the switch to use of DNA for genetic storage (Poole and Logan 2005). Together, these results point to a very late origin for DNA, post-dating the diversification of the primary lineages from the LUCA (Fig. 2a).

It is worth noting that, while the primary focus in thinking about the evolution of the genetic material is often on fidelity of replication, in light of the fact of RNA repair, an alternative model for the origin of DNA is at least as plausible, particularly in light of horizontal gene transfer of both replicative and synthetic enzymes. That model, proposed by Forterre, is that the enzymes required for synthesis and replication initially evolved in viruses. In this model, viral-host cell coevolution drove modification of viral genetic material, much as it does in modern viruses (Forterre 2002, 2006). These modifications served to provide viral genome protection from cellular defences (RNases in the first instance), and these host-virus interactions subsequently led to viral-to-cellular transfers, neutralising the viral advantage. Forterre's model helps to account for the wide variety of enzymes that perform the same functions in deoxyribonucleotide synthesis and DNA replication (Forterre 2013), something that other models fail to comprehensively address.

A Bottom-Up Perspective on the Origin of DNA

The reverse deoxyriboaldolase (DERA) reaction (Fig. 3) (Racker 1951, 1952) has been noted by prebiotic chemists

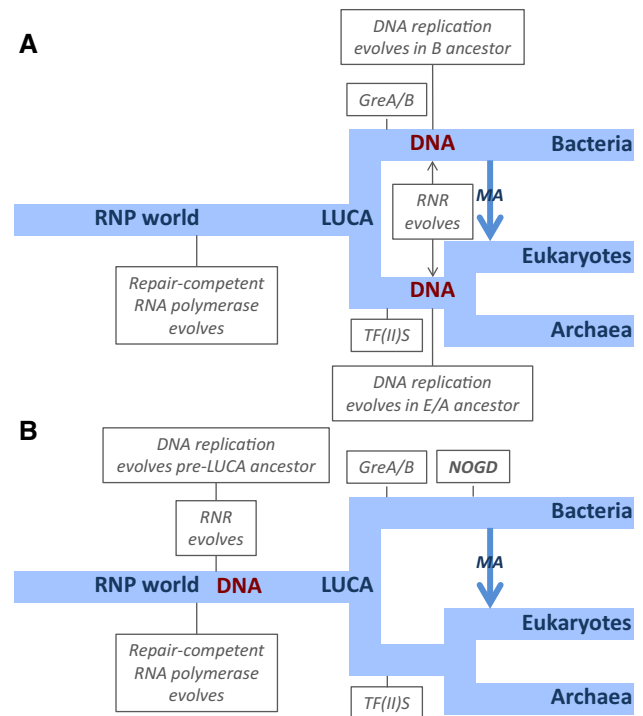


Fig. 2 Two interpretations of the biological data on DNA origins. The observation that much of the core machinery for DNA replication differs between Bacteria on the one hand, and Archaea/Eukaryotes on the other (Forterre et al. 2004; Leipe et al. 1999) can be interpreted in two different ways (Poole and Logan 2005). **a** The case for an RNA-based LUCA can be made under a model where a repair-competent RNA polymerase is utilised in a replicative capacity, with lineage-independent origins for RNA proofreading fidelity factors (the GreA/B system in Bacteria, and the TF(II)S system in Archaea/Eukaryotes). Later, the replicative DNA apparatus emerges twice independently (Leipe et al. 1999). Under this model, the origin of ribonucleotide reduction is late but is obscured by subsequent horizontal gene transfer events (Lundin et al. 2010). **b** The case for a DNA-based LUCA can be made on the exact same data by invoking non-orthologous gene displacement (NOGD) (Leipe et al. 1999). For clarity, the bacterial lineage is depicted as being subject to NOGD, but this could equally have occurred in the Archaeal/Eukaryotic lineage. The key point of this model is that one of the two apparatuses is evolutionarily older than the other, with the displacement possibly occurring through acquisition of viral replication genes (Forterre 2013). This model fits with the observation that some features of the DNA replication machinery appear to be ubiquitous as do some parts of the DNA repair machinery. It also permits placement of ribonucleotide reduction before the divergence of the three domains (while still accepting modern transfer events obscure the deep origins of ribonucleotide reductase). It is also possible for RNA polymerase fidelity factors to be placed pre-LUCA under non-orthologous gene displacement (Poole and Logan 2005). *MA* mitochondrial ancestor. Note that there is ongoing debate regarding the exact relationship between eukaryotes and the Archaea (Poole and Gribaldo 2014). Each domain is shown as a distinct group for clarity, though under an archaeal origin for eukaryotes the interpretation of this figure is unchanged

(Benner et al. 1989; Burton and Lehman 2009) as a chemically far simpler—and therefore possibly much earlier—route to deoxyribonucleotide synthesis. DERA is

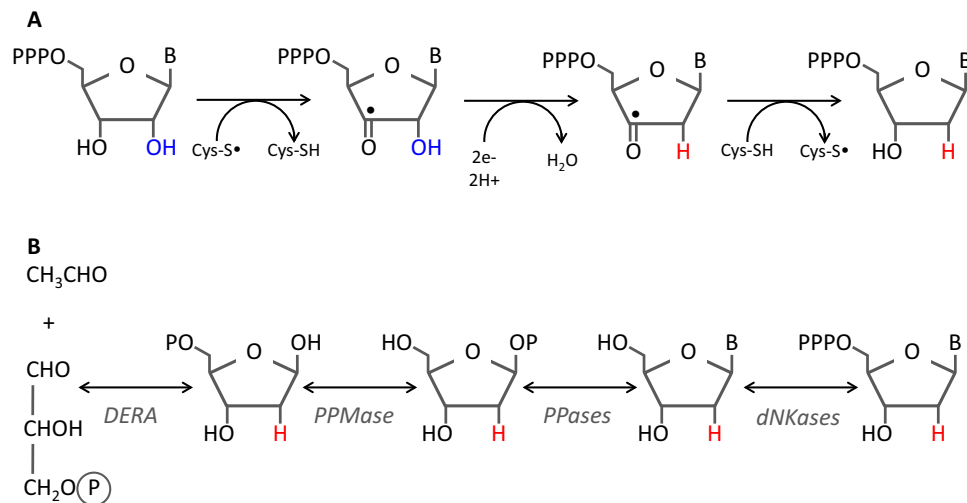


Fig. 3 The chemistry of ribonucleotide reduction and deoxyribonucleotide salvage. **a** Generalised pathway for ribonucleotide reduction, showing steps common to all ribonucleotide reductase classes. Production of a stable cysteinyl radical in the active site of ribonucleotide reductase occurs in all variants of ribonucleotide reduction, though the initial formation of the radical differs across classes. The radical is transferred to the ribonucleotide substrate (depicted here as an NTP, though this can be either NDP or NTP, depending on the organism), driving protonation of the 2' hydroxyl group, with the release of a water molecule. This step involves a separate reductant, and the nature of the reductant differs between

ribonucleotide reductase classes. Finally, the cysteinyl radical is regenerated before being returned to the original site of radical generation (cofactor or a different protein residue). See (Hofer et al. 2012) for a recent review of ribonucleotide reduction. **b** Deoxyribonucleosides can be produced from glyceraldehyde-3-phosphate and acetaldehyde via the reverse of the deoxyriboaldolase (DERA) reaction, forming 2-deoxyribose-5-phosphate. The subsequent action of phosphopentomutase (PPMase), phosphorylases (PPases) and deoxynucleotide kinases (dNKases) generates dNTPs. Modern enzymes can drive the synthesis of deoxyribonucleosides but have thus far only been shown to do so in vitro (Horinouchi et al. 2006c)

involved in modern salvage, through which deoxyribonucleosides are catabolised to acetaldehyde and glyceraldehyde-3-phosphate, where they can be incorporated into central metabolism. However, prior to the discovery of ribonucleotide reduction (Reichard and Rutberg 1960), the reverse DERA pathway was considered the most likely cellular route for deoxyribonucleotide synthesis (Racker 1951, 1952). That said, owing to low intracellular levels of acetaldehyde, the reaction is only known to proceed in the degradative direction in vivo, and synthesis is only observed in cell extracts (Horinouchi et al. 2006c; Racker 1952).

As elegantly laid out by Burton and Lehman (Burton and Lehman 2009), all the steps in the reaction outlined in Fig. 3b could plausibly be performed by ribozyme chemistry, suggesting a bottom-up approach to this question could bear experimental fruit. In contrast, to be within reach of ribozyme chemistry, ribonucleotide reduction would need to proceed via an alternative chemical route (Burton and Lehman 2009) as the chemistry employed by modern RNRs (Fig. 3a) remains, on current knowledge, beyond the reach of ribozymes (Poole et al. 2000). Thus, if DNA usage did evolve early, the ribozyme-based DERA pathway currently provides the most plausible chemical route.

Building a Case for the Early Origin of DNA via the DERA Pathway

An important goal for origins research is to find connections between emerging biological and chemical insights. At first glance, the data underlying the top-down model, where DNA evolved late (Fig. 2a), appear incompatible with an early origin for DNA via reverse DERA. However, as with all biological reconstructions of the deep past, these events are clouded by uncertainty, and more than one interpretation is possible. Indeed, by invoking non-orthologous gene displacement events (Leipe et al. 1999), the biological data can be shown to be equally compatible with an earlier origin for DNA, and it is in principle possible to place DNA in the LUCA (Forterre 2013; Poole 2011; Poole and Logan 2005) (Fig. 2b).

That comparative analysis as a general approach is limited in its capacity to produce a definitive picture is well established. Reconstructions of the genomic content of the LUCA show that, beyond the ribosome, a smattering of ribosome-associated processes and the conserved core of multisubunit RNA polymerases, it is difficult to consistently trace many more features to the LUCA (Goldman et al. 2013; Harris et al. 2003; Hoepfner et al. 2012; Koonin 2003). These processes are clearly insufficient to

run a cell, and there is widespread acceptance that horizontal gene transfer and non-orthologous gene displacement must have played a role in the shaping of modern biology (Koonin 2003; Nelson-Sathi et al. 2014; Poole 2009; Vetsigian et al. 2006; Woese 2002). Loss of information is an equally large problem for reconstructing past states, both in terms of gene losses, which risk being interpreted as evidence for late emergence (Becerra et al. 1997; Glansdorff et al. 2008), and exponential loss of signal in sequence data, which limits phylogenetic reconstruction based on sequence data (Penny and Zhong 2014) (though structural information can help here (Daly et al. 2013; Lundin et al. 2012).

In light of these processes, the LUCA could well have carried a full set of DNA replication machinery, and in this regard it is noteworthy that key parts of the DNA replication machinery do appear universal. The clamp-loader and clamp, which give DNA polymerases their processivity, are universally conserved (Kelch et al. 2012; Leipe et al. 1999), as is RNase HIII, which facilitates removal of RNA primers used during replication (Brinddefalk et al. 2013; Tadokoro and Kanaya 2009). Clamp-loader, clamp and RNase H genes are also found in viruses, consistent with the viral-cellular coevolution model for DNA origins (Forterre 2013). The recent discovery that the universally conserved replicative helicase, UvrD, is intimately involved in recruiting the DNA repair machinery (some of which has also been argued to trace to the LUCA (Eisen and Hanawalt 1999)) at the site of active transcription (Epshtein et al. 2014), is also compatible with a DNA-based LUCA. Moreover, that RNRs, thymidylate synthases, and even DNA polymerases, are subject to ongoing horizontal gene transfer and non-orthologous gene displacement, including between viruses and their hosts, suggests that the machinery involved in deoxyribonucleotide synthesis and DNA replication cannot be interpreted exclusively in terms of late gains.

Finally, support for an early origin of DNA has also emerged via modelling. In simulation studies, it was shown that a DNA-like capacity is advantageous early because it eliminates the trade-off between the information storage and catalytic functions of RNA, and in fact makes the modelled system more robust to invasion by parasites (Takeuchi et al. 2011).

If the interpretation in Fig. 2b is accepted, then the mutual incompatibility of a very early origin for DNA and comparative analyses vanishes. Such an interpretation removes the difficulty of placing DNA in the LUCA, though it does not enable us to directly reconstruct the replicative machinery of this early stage.

Plausibility is important, but biologists are equally concerned with historical signal. Significantly, RNRs are ubiquitous (bar a handful of bacteria, which lost RNRs

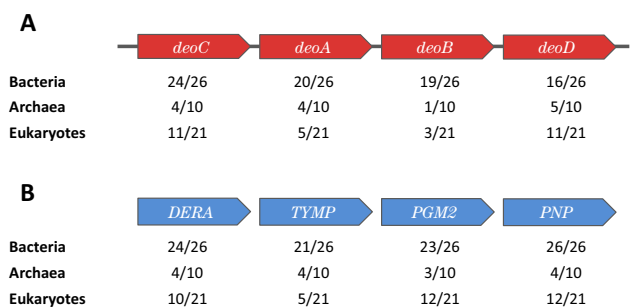


Fig. 4 The deoxyriboaldolase pathway is present in all three domains. **a** The proteins encoded by the *E. coli deo* operon were used as query sequences for a pHMMER search (Finn et al. 2011) against the nr protein database. Red arrows indicate ORFs and their position in the *E. coli deo* operon. The number of phyla in each domain with significant hits (E -value < 0.01) were counted (see Tables S1–3 for phylum-level results). **b** Genes from the *H. sapiens* genome with identical function to those of the *E. coli deo* operon are shown in blue. The proteins encoded by these genes were used as query sequences for a pHMMER search against the nr protein database. The number of phyla in each domain with significant hits (E -value < 0.01) were counted (see Tables S1–3 for phylum-level results)

during adaptation to an obligate intracellular lifestyle and now derive deoxyribonucleotides from their host) (Lundin et al. 2010, 2009). Thus, if the reverse DERA reaction ever was used for deoxyribonucleotide synthesis, it has long ceased to perform this function. That said, salvage via DERA is widespread, and the key genes required for operation of this pathway can be readily identified in all three domains (Fig. 4).

Towards Testing Deep Evolution in a Cellular, Experimental Context

If DNA synthesis and replication are placed in the LUCA, could we take the next step and consider the reverse deoxyriboaldolase reaction as a possible early route to DNA? This has yet to be investigated. Certainly, the distribution of the enzymes is compatible with an ancient origin (Fig. 4), but more crucially, biology is now developing tools (Elena and Lenski 2003; Gibson et al. 2010; Lutz and Patrick 2004) to enable hypothesis testing on a level that matches the incredible success of SELEX experiments in expanding the chemistry of the RNA world (Breaker and Joyce 2014; Yarus 1999).

There are several important ways that biology may help connect to chemical research into DNA origins. We lay these out as a series of general questions that could be applied to the empirical study of any pathway:

1. Can the product be produced by existing enzymes?
2. Can the product be produced in vivo by existing enzymes?

3. Can the reaction replace the incumbent pathway?

For DNA origins, the answer to the first question is straightforward: the reverse DERA reaction was known to operate at the time that deoxyriboaldolase was first characterised (Racker 1951, 1952). That said, there is an important parallel between SELEX experiments to test the feasibility of the individual steps in this pathway in a hypothetical RNA world (Burton and Lehman 2009) and microbial process engineering. Here, screening has yielded novel natural versions of these enzymes that, together, can improve the synthetic yield of the desired product (Horinouchi et al. 2006a, b, c, 2003, 2012; Ogawa et al. 2003). Thus, pathway optimisation may be rapidly achieved through creation of a patchwork of enzymes from different species.

SELEX and directed evolution of proteins test one enzyme or reaction at a time. The patchwork approach thus helps test more complex processes in a cell-free environment, such as deoxyribonucleoside synthesis (Horinouchi et al. 2006b), and even simplified 19-amino acid genetic codes (Kawahara-Kobayashi et al. 2012). These examples address Question 1 above in that they show that a process can be carried out by the components of biological systems.

In the case of undertaking replacement of one reaction by another, the solutions to questions 2 and 3 above are likely to be linked. Question 2 is vitally important because it asks whether a reaction will work in the complex context of a cell. In the case of deoxyribonucleotide synthesis, this permits us to address the original critique of reverse deoxyriboaldolase as a route to deoxyribonucleosides: that there would be insufficient starting substrate to ever drive the reaction in the direction of synthesis. A complication with an essential process such as deoxyribonucleotide synthesis is that knocking out the genes for the existing pathway (Question 3) can only be successful in the context of success in production by an alternative route (Question 2). Certainly, under normal circumstances, it is not possible to knock out all RNRs in *E. coli* simultaneously because this pathway is essential, even under conditions where a bespoke deoxyriboaldolase operon is overexpressed (DS, NH, RJC personal observations). Consequently, the challenge in testing the biological viability of the reverse DERA pathway is finding conditions that should favour synthesis.

Analysis of the DERA pathway suggests the key problem is acetaldehyde. Acetaldehyde is highly reactive and may not be available in vivo in sufficient amounts for synthesis. Moreover, the levels of acetaldehyde used in microbial process engineering for deoxyribonucleoside synthesis are toxic to *E. coli* (Horinouchi et al. 2006a; Ogawa et al. 2003). Additionally, acetaldehyde is

important for NAD⁺ regeneration. *E. coli* carries multiple alcohol dehydrogenase genes, which reduce acetaldehyde to ethanol, oxidising NADH in the process. Thus, if acetaldehyde is diverted from NAD⁺ production, this may upset NAD⁺/NADH pools. All of these hurdles would need to be negotiated in testing the possibility of the DERA pathway supporting deoxyribonucleotide synthesis in a cellular system.

Rapid Takeover of Deoxyribonucleotide Synthesis by Ribonucleotide Reduction

There are good reasons to expect that, if early cells did utilise deoxyriboaldolase for deoxyribonucleotide synthesis, this pathway was unlikely to persist as a primary synthetic route following the advent of ribonucleotide reduction. We will first consider the pros of modern ribonucleotide reduction, before comparing this to the DERA pathway.

Modern deoxyribonucleotide pools are one to two orders of magnitude lower than nucleotide pools (Nick McElhinny et al. 2010), so diverting a small fraction of the latter to deoxyribonucleotide synthesis following advent of ribonucleotide reduction likely caused minimal metabolic disruption to the cell. Ribonucleotide reduction is an irreversible reaction, and RNRs exhibit sophisticated allosteric regulation (as does pyruvate formate lyase, which is homologous to RNRs, and appears evolutionarily closest to class III RNRs (Leppanen et al. 1999; Logan et al. 1999)). Hence, modern deoxyribonucleotide synthesis can be tightly controlled, thereby avoiding unbalanced or elevated dNTP levels, which would lead to increased mutation rates (Hofer et al. 2012).

In contrast to ribonucleotide reduction, the DERA reaction is reversible. Therefore, there is potential for futile cycling. If deoxyribonucleotides were initially synthesised via the DERA pathway, deoxyribonucleotide pools may have fluctuated based on whether substrates (acetaldehyde and glyceraldehyde-3-phosphate) or product (deoxyribonucleosides) were in abundance (Horinouchi et al. 2006c; Ogawa et al. 2003). Given the toxicity of acetaldehyde—which, in modern cells causes DNA damage, formation of protein adducts, and free radical production (Dellarco 1988)—2-deoxyribose-5-phosphate production may have initially provided a means of eliminating this toxic molecule. However, without modification of 2-deoxyribose-5-phosphate, the resulting low levels of acetaldehyde could drive breakdown of deoxyribonucleosides. Downstream reactions (Fig. 3) could have favoured deoxyribonucleotide synthesis, and reduced futile cycling, but the entire reaction is still reversible, which would have resulted in poor control of deoxyribonucleotide pools, a phenomenon that is

known in modern systems to lead directly to mutational pressures (Yao et al. 2013). Finally, if the DERA pathway operated in the direction of synthesis, there may have been competition between DERA-driven deoxyribonucleotide synthesis on the one hand, and use of acetaldehyde as an electron acceptor for NAD⁺ regeneration (producing ethanol) on the other.

In contrast, ribonucleotide reduction would have offered efficient, irreversible deoxyribonucleotide formation, which would have been critical to reducing futile cycling, and would have provided better control of deoxyribonucleotide synthesis. Utilising a small fraction of the ribonucleotide pool would have had minimal disruption on existing roles for this substrate, and permitted more effective NAD⁺ regeneration through fermentation. Thus, if the DERA pathway did evolve much earlier than ribonucleotide reduction, comparison of the two pathways suggests DERA would be completely superseded as a synthetic pathway following the advent of ribonucleotide reduction.

Concluding Remarks

There has been enormous progress in understanding the evolutionary origins of DNA. However, different views have been reached based on assessment of the evolutionary history of the cellular apparatus for deoxyribonucleotide synthesis and DNA replication (Forterre 2002; Leipe et al. 1999; Poole and Logan 2005) or from chemical considerations (Burton and Lehman 2009). One interpretation of the biological data is that DNA may have evolved after Bacteria diverged from Archaea and Eukaryotes, such that the LUCA possessed an RNA genome. However, the data could also be consistent with a DNA-based LUCA, but with DNA evolving after the advent of complex protein-based catalysts. In contrast, if chemical routes to deoxyribonucleotides are not limited to ribonucleotide reduction, then DNA could have emerged in an RNA world, before templated protein synthesis. While these views appear diametrically opposed, it is not clear that these they are actually incompatible (Poole 2011). Indeed, we see no reason that an early RNA world origin of deoxyribonucleotide synthesis cannot be compatible with a much later origin of the modern enzymes for deoxyribonucleotide synthesis. A promising avenue to address this is to establish whether it is possible to create the conditions wherein a cell can synthesize deoxyribonucleotides via acetaldehyde and glyceraldehyde-3-phosphate instead of via ribonucleotide reduction. It is clear that modern DERA pathway enzymes can indeed drive deoxyribonucleoside synthesis (Horinouchi et al. 2006a), and this shows great promise for industrial-scale production of deoxyribonucleosides (Horinouchi

et al. 2012). From an evolutionary perspective, assessment of the capacity for this pathway to replace ribonucleotide reduction in vivo is the crucial next step.

Acknowledgments We thank two anonymous reviewers for helpful comments on the manuscript. JO and AMP gratefully acknowledge receipt of a Japan Society for the Promotion of Science (JSPS) Invitation Fellowship (S137 22), plus research funding via the New Zealand-Japan Joint Research Project Programme administered by JSPS and the Royal Society of New Zealand. AMP gratefully acknowledges the support of the Royal Society of New Zealand via a Rutherford Discovery Fellowship.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Anastasi C, Buchet FF, Crowe MA, Parkes AL, Powner MW, Smith JM, Sutherland JD (2007) RNA: prebiotic product, or biotic invention? *Chem Biodivers* 4:721
- Augustin MA, Huber R, Kaiser JT (2001) Crystal structure of a DNA-dependent RNA polymerase (DNA primase). *Nat Struct Biol* 8:57
- Bailey S, Wing RA, Steitz TA (2006) The structure of *T. aquaticus* DNA polymerase III is distinct from eukaryotic replicative DNA polymerases. *Cell* 126:893
- Becerra A, Islas S, Leguina JI, Silva E, Lazcano A (1997) Polyphyletic gene losses can bias backtrack characterizations of the cenancestor. *J Mol Evol* 45:115
- Benner SA, Ellington AD, Tauer A (1989) Modern metabolism as a palimpsest of the RNA world. *Proc Natl Acad Sci USA* 86:7054
- Breaker RR, Joyce GF (2014) The expanding view of RNA and DNA function. *Chem Biol* 21:1059
- Brindefalk B, Dessailly BH, Yeats C, Orengo C, Werner F, Poole AM (2013) Evolutionary history of the TBP-domain superfamily. *Nucleic Acids Res* 41:2832
- Burton AS, Lehman N (2009) DNA before proteins? Recent discoveries in nucleic acid catalysis strengthen the case. *Astrobiology* 9:125
- Collins LJ, Moulton V, Penny D (2000) Use of RNA secondary structure for studying the evolution of RNase P and RNase MRP. *J Mol Evol* 51:194
- Cramer P (2002) Multisubunit RNA polymerases. *Curr Opin Struct Biol* 12:89
- Daly TK, Sutherland-Smith AJ, Penny D (2013) Beyond BLASTing: tertiary and quaternary structure analysis helps identify major vault proteins. *Genome Biol Evol* 5:217
- del Mar Cendra M, Juarez A, Torrents E (2012) Biofilm modifies expression of ribonucleotide reductase genes in *Escherichia coli*. *PLoS One* 7:e46350
- Dellarco VL (1988) A mutagenicity assessment of acetaldehyde. *Mutat Res* 195:1
- Dwivedi B, Xue B, Lundin D, Edwards RA, Breitbart M (2013) A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. *BMC Evol Biol* 13:33
- Dworkin JP, Lazcano A, Miller SL (2003) The roads to and from the RNA world. *J Theor Biol* 222:127
- Eisen JA, Hanawalt PC (1999) A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* 435:171

- Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* 4:457
- Epshtein V, Kamarthapu V, McGary K, Svetlov V, Ueberheide B, Proshkin S, Mironov A, Nudler E (2014) UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* 505:372
- Fels A, Hu K, Riesner D (2001) Transcription of potato spindle tuber viroid by RNA polymerase II starts predominantly at two specific sites. *Nucleic Acids Res* 29:4589
- Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39:W29
- Forterre P (1999) Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Mol Microbiol* 33:457
- Forterre P (2002) The origin of DNA genomes and DNA replication proteins. *Curr Opin Microbiol* 5:525
- Forterre P (2006) Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc Natl Acad Sci USA* 103:3669
- Forterre P (2013) Why are there so many diverse replication machineries? *J Mol Biol* 425:4714
- Forterre P, Filée J, Myllykallio H (2004) Origin and evolution of DNA and DNA replication machineries. In: de Pouplana LR (ed) *The genetic code and the origin of life*. Landes Bioscience, Georgetown, TX, pp 145–168
- Freeland SJ, Knight RD, Landweber LF (1999) Do proteins predate DNA? *Science* 286:690
- Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329:52
- Gilbert W (1986) The RNA world. *Nature* 319:618
- Glandsdorff N, Xu Y, Labedan B (2008) The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol Direct* 3:29
- Goldman AD, Bernhard TM, Dolzhenko E, Landweber LF (2013) LUCApedia: a database for the study of ancient life. *Nucleic Acids Res* 41:D1079
- Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35:849
- Harris JK, Kelley ST, Spiegelman GB, Pace NR (2003) The genetic core of the Universal Ancestor. *Genome Res* 13:407
- Hoepfner MP, Gardner PP, Poole AM (2012) Comparative analysis of RNA families reveals distinct repertoires for each domain of life. *PLoS Comput Biol* 8:e1002752
- Hofer A, Crona M, Logan DT, Sjöberg BM (2012) DNA building blocks: keeping control of manufacture. *Crit Rev Biochem Mol Biol* 47:50
- Horinouchi N, Ogawa J, Sakai T, Kawano T, Matsumoto S, Sasaki M, Mikami Y, Shimizu S (2003) Construction of deoxyriboaldolase-overexpressing *Escherichia coli* and its application to 2-deoxyribose 5-phosphate synthesis from glucose and acetaldehyde for 2'-deoxyribonucleoside production. *Appl Environ Microbiol* 69:3791
- Horinouchi N, Ogawa J, Kawano T, Sakai T, Saito K, Matsumoto S, Sasaki M, Mikami Y, Shimizu S (2006a) Biochemical retrosynthesis of 2'-deoxyribonucleosides from glucose, acetaldehyde, and a nucleobase. *Appl Microbiol Biotechnol* 71:615
- Horinouchi N, Ogawa J, Kawano T, Sakai T, Saito K, Matsumoto S, Sasaki M, Mikami Y, Shimizu S (2006b) Efficient production of 2-deoxyribose 5-phosphate from glucose and acetaldehyde by coupling of the alcoholic fermentation system of Baker's yeast and deoxyriboaldolase-expressing *Escherichia coli*. *Biosci Biotechnol Biochem* 70:1371
- Horinouchi N, Ogawa J, Kawano T, Sakai T, Saito K, Matsumoto S, Sasaki M, Mikami Y, Shimizu S (2006c) One-pot microbial synthesis of 2'-deoxyribonucleoside from glucose, acetaldehyde, and a nucleobase. *Biotechnol Lett* 28:877
- Horinouchi N, Sakai T, Kawano T, Matsumoto S, Sasaki M, Hibi M, Shima J, Shimizu S, Ogawa J (2012) Construction of microbial platform for an energy-requiring bioprocess: practical 2'-deoxyribonucleoside production involving a C–C coupling reaction with high energy substrates. *Microb Cell Fact* 11:82
- Hud NV, Cafferty BJ, Krishnamurthy R, Williams LD (2013) The origin of RNA and “my grandfather's axe”. *Chem Biol* 20:466
- Jeffares DC, Poole AM, Penny D (1998) Relics from the RNA world. *J Mol Evol* 46:18
- Kawahara-Kobayashi A, Masuda A, Arais Y, Sakai Y, Kohda A, Uchiyama M, Asami S, Matsuda T, Ishitani R, Dohmae N, Yokoyama S, Kigawa T, Nureki O, Kiga D (2012) Simplification of the genetic code: restricted diversity of genetically encoded amino acids. *Nucleic Acids Res* 40:10576
- Kelch BA, Makino DL, O'Donnell M, Kuriyan J (2012) Clamp loader ATPases and the evolution of DNA replication machinery. *BMC Biol* 10:34
- Koonin EV (2003) Comparative genomics, minimal gene-sets and the Last Universal Common Ancestor. *Nat Rev Microbiol* 1:127
- Kun A, Santos M, Szathmary E (2005) Real ribozymes suggest a relaxed error threshold. *Nat Genet* 37:1008
- Lai MM (2005) RNA replication without RNA-dependent RNA polymerase: surprises from Hepatitis Delta Virus. *J Virol* 79:7951
- Larsson KM, Jordan A, Eliasson R, Reichard P, Logan DT, Nordlund P (2004) Structural mechanism of allosteric substrate specificity regulation in a ribonucleotide reductase. *Nat Struct Mol Biol* 11:1142
- Lazcano A, Guerrero R, Margulis L, Oro J (1988) The evolutionary transition from RNA to DNA in early cells. *J Mol Evol* 27:283
- Lehman N (2003) A case for the extreme antiquity of recombination. *J Mol Evol* 56:770
- Lehmann E, Brueckner F, Cramer P (2007) Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* 450:445
- Leipe DD, Aravind L, Koonin EV (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res* 27:3389
- Leppanen VM, Merckel MC, Ollis DL, Wong KK, Kozarich JW, Goldman A (1999) Pyruvate formate lyase is structurally homologous to type I ribonucleotide reductase. *Structure* 7:733
- Logan DT, Andersson J, Sjöberg BM, Nordlund P (1999) A glycol radical site in the crystal structure of a class III ribonucleotide reductase. *Science* 283:1499
- Lundin D, Torrents E, Poole AM, Sjöberg BM (2009) RNRdb, a curated database of the universal enzyme family ribonucleotide reductase, reveals a high level of misannotation in sequences deposited to Genbank. *BMC Genom* 10:589
- Lundin D, Gribaldo S, Torrents E, Sjöberg BM, Poole AM (2010) Ribonucleotide reduction: horizontal transfer of a required function spans all three domains. *BMC Evol Biol* 10:383
- Lundin D, Poole AM, Sjöberg BM, Högbon M (2012) Use of structural phylogenetic networks for classification of the ferritin-like superfamily. *J Biol Chem* 287:20565
- Lutz S, Patrick WM (2004) Novel methods for directed evolution of enzymes: quality, not quantity. *Curr Opin Biotechnol* 15:291
- Myllykallio H, Lipowski G, Leduc D, Filée J, Forterre P, Liebl U (2002) An alternative flavin-dependent mechanism for thymidylate synthesis. *Science* 297:105
- Nelson-Sathi S, Sousa FL, Roettger M, Lozada-Chavez N, Thiergart T, Janssen A, Bryant D, Landan G, Schonheit P, Siebers B,

- McInerney JO, Martin WF (2014) Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature*. doi:10.1038/nature13805
- Nick McElhinny SA, Watts BE, Kumar D, Watt DL, Lundstrom EB, Burgers PM, Johansson E, Chabes A, Kunkel TA (2010) Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci USA* 107:4949
- Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* 289:920
- Noller HF (2010) Evolution of protein synthesis from an RNA World. *Cold Spring Harb Perspect Biol* 4:a003681
- Ogawa J, Saito K, Sakai T, Horinouchi N, Kawano T, Matsumoto S, Sasaki M, Mikami Y, Shimizu S (2003) Microbial production of 2-deoxyribose 5-phosphate from acetaldehyde and triosephosphate for the synthesis of 2'-deoxyribonucleosides. *Biosci Biotechnol Biochem* 67:933
- Penny D, Zhong B (2014) Two fundamental questions about protein evolution. *Biochimie*. doi:10.1016/j.biochi.2014.10.020
- Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, Stepanov VG, Harvey SC, Fox GE, Wartell RM, Hud NV, Williams LD (2014) Evolution of the ribosome at atomic resolution. *Proc Natl Acad Sci USA* 111:10251
- Poole AM (2006) Getting from an RNA world to modern cells just got a little easier. *BioEssays* 28:105
- Poole AM (2009) Horizontal gene transfer and the earliest stages of the evolution of life. *Res Microbiol* 160:473
- Poole AM (2011) On alternative biological scenarios for the evolutionary transitions to DNA and biological protein synthesis. In: Egel R, Lankenau D-H, Mulikidjanian AY (eds) *Origins of life: the primal self-organization*. Springer, Berlin, pp 209–223
- Poole AM, Gribaldo S (2014) Eukaryotic Origins: How and when was the mitochondrion acquired? *Cold Spring Harb Perspect Biol*. doi:10.1101/cshperspect.a015990
- Poole AM, Logan DT (2005) Modern mRNA proofreading and repair: clues that the Last Universal Common Ancestor possessed an RNA genome? *Mol Biol Evol* 22:1444
- Poole A, Penny D, Sjöberg B-M (2000) Methyl-RNA: an evolutionary bridge between RNA and DNA? *Chem Biol* 7:R207
- Poole A, Penny D, Sjöberg BM (2001) Confounded cytosine! Tinkering and the evolution of DNA. *Nat Rev Mol Cell Biol* 2:147
- Poole AM, Logan DT, Sjöberg B-M (2002) The evolution of the ribonucleotide reductases: much ado about oxygen. *J Mol Evol* 55:180
- Racker E (1951) Enzymatic synthesis of deoxypentose phosphate. *Nature* 167:408
- Racker E (1952) Enzymatic synthesis and breakdown of desoxyribose phosphate. *J Biol Chem* 196:347
- Reaney DC (1987) Genetic error and genome design. *Cold Spring Harb Symp Quant Biol* 52:751
- Reichard P (1993) From RNA to DNA, why so many ribonucleotide reductases? *Science* 260:1773
- Reichard P, Rutberg L (1960) Formation of deoxycytidine 5'-phosphate from cytidine 5'-phosphate with enzymes from *Escherichia coli*. *Biochim Biophys Acta* 37:554
- Sintchak MD, Arjara G, Kellogg BA, Stubbe J, Drennan CL (2002) The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer. *Nat Struct Biol* 9:293
- Tadokoro T, Kanaya S (2009) Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J* 276:1482
- Takeuchi N, Hogeweg P, Koonin EV (2011) On the origin of DNA genomes: evolution of the division of labor between template and catalyst in model replicator systems. *PLoS Comput Biol* 7:e1002024
- Tauer A, Benner SA (1997) The B12-dependent ribonucleotide reductase from the archaeobacterium *Thermoplasma acidiphila*: an evolutionary solution to the ribonucleotide reductase conundrum. *Proc Natl Acad Sci USA* 94:53
- Torrents E, Aloy P, Gibert I, Rodriguez-Trelles F (2002) Ribonucleotide reductases: divergent evolution of an ancient enzyme. *J Mol Evol* 55:138
- Tourasse NJ, Gouy M (1999) Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol Phylogenet Evol* 13:159
- Uhlin U, Eklund H (1994) Structure of ribonucleotide reductase protein R1. *Nature* 370:533
- Vetsigian K, Woese C, Goldenfeld N (2006) Collective evolution and the genetic code. *Proc Natl Acad Sci USA* 103:10696
- Werner F, Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* 9:85
- White HB 3rd (1976) Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* 7:101
- Woese CR (2002) On the evolution of cells. *Proc Natl Acad Sci USA* 99:8742
- Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 74:5088
- Yao NY, Schroeder JW, Yurieva O, Simmons LA, O'Donnell ME (2013) Cost of rNTP/dNTP pool imbalance at the replication fork. *Proc Natl Acad Sci USA* 110:12942
- Yarus M (1999) Boundaries for an RNA world. *Curr Opin Chem Biol* 3:260
- Yarus M (2002) Primordial genetics: phenotype of the ribocyte. *Annu Rev Genet* 36:125
- Yarus M (2011) Getting past the RNA world: the initial Darwinian ancestor. *Cold Spring Harb Perspect Biol* 3:a003590
- Zenkin N, Yuzenkova Y, Severinov K (2006) Transcript-assisted transcriptional proofreading. *Science* 313:518